A genome wide analysis of key genes controlling diastatic power activity in UK barley (DPGENES)

Hoad, SP; Joanne Russell; Mark Looseley; Micha Bayer; Hazel Bull; Luke Ramsay; William Thomas; Allan Booth; Jenny Morris; Pete Hedley; Linde Hess; James Brosnan

Print publication: 01/10/2017

September 2017

Project Report No. PR583

A genome wide analysis of key genes controlling diastatic power activity in UK barley (DPGENES)

Joanne Russell[1], Mark Looseley[1*], Micha Bayer[1], Hazel Bull[1], Luke Ramsay[1], William Thomas[1], Allan Booth[1], Jenny Morris[1], Pete Hedley[1], Steve Hoad[2], Linde Hess[3] and James Brosnan[3]

[1]Cell and Molecular Sciences, The James Hutton Institute, Dundee, Scotland
[2]SRUC, Peter Wilson Building, Kings Buildings, West Mains Road, Edinburgh, EH9 3JG
[3]Scotch Whisky Research Institute, Research Avenue North, Riccarton, Edinburgh EH14 4AP

This is the final report of a 27-month project (21130001) which started in April 2015. The work was funded by AHDB and a contract for £207,967 from AHDB Cereals & Oilseeds.

CONTENTS

1. Abstract

Diastatic Power (DP) is an important quality trait for malt used in adjunct brewing and distilling. Substantial genetic variation for DP exists within UK elite barley cultivars, but breeding progress has been slow due to the limited demand, compared to the overall barley market, and difficulties in assessing DP. The aim of this project was to develop a suite of genetic markers and robust phenotypic screening methods to identify differences in barley malt diastase activity that can be used by breeders, testing authorities, and maltsters for the early identification and promotion of new varieties for the grain distilling market. To do this, we first identified subsets of barley accessions that contrast for diastase activity using pre-existing data for both spring and winter germplasm. These subsets were the basis for marker development (A) and phenotypic screening protocols (B).

(A) Twelve high DP and twelve low DP pools of DNA for spring and winter lines were sequenced using a gene enrichment approach, generating over 84,000 polymorphic SNP markers. Using allele frequency differences between low and high DP pools, we identified 66 and 32 SNPs that distinguished between high and low DP in the winters and springs, respectively. For each chromosome region, we chose the marker that showed the highest differentiation between pools, several of these were collocated with known diastase genes, or genes which were annotated with a putative diastase function. Eight KASP marker assays (marker favoured by UK and European breeders) were designed from the SNPs identified from the winters and five for springs. These were tested on a large set of winter and spring germplasm, and a strong correlation ($r = 0.92$) between genotype and true DP was observed. Further validation was carried out using independent germplasm, supplied by UK and European breeders, which were genotyped, DP predicted and confirmed where possible by micro-malting at Scotch Whisky Research Institute (SWRI).

(B) The subsets selected were sown under a range of conditions expected to differentiate between high and low DP lines. Winter and spring lines were trialled under both standard and high nitrogen fertiliser treatments, and yield data showed little difference between high and low DP varieties, but in contrast grain nitrogen differed between subsets considerably, at both rates of nitrogen. Based on these initially findings an improved assessment protocol to complement the markers in an integrated screening package enable breeders to produce new varieties that are specifically targeted at the production of high DP malt.

2. Introduction

One of the most economically significant uses of barley is in the production of alcohol following malting, with about 20% of annual UK barley production being used for processing, the vast majority of which is used by maltsters. During the malting process, endogenous proteolytic and amylolytic enzymes (either present within the mature grain or generated during germination) are released and modify the barley endosperm, with starch being converted into fermentable sugars. The breakdown of starch is catalysed by a group of enzymes known as diastases and their combined ability to do so is called diastatic power (DP). High DP is a key attribute in the purchase of malting barley to be used to digest cooked starch adjunct in grain distilling and adjunct brewing. There is a known positive correlation between diastase activity and grain nitrogen content, so maltsters typically purchase higher nitrogen grain lots to supply the needs of grain distillers. This now represents a significant market, and breeding opportunities exist for new varieties that combine the high agronomic performance associated with modern varieties and high levels of diastase activity.

Progress in breeding for high diastase has stagnated in the UK because the character is generally only measured under the lower nitrogen regimes more typical of malting barley and the market demand is relatively small (<10% of total UK malting barley purchases). In addition, whilst limit dextrinase, alpha-amylase and beta-amylase are known to be the key enzymes affecting diastase activity, repeated selection for good malting quality means there is little variation for the structural genes for these enzymes amongst the current UK elite barley genepool. Nevertheless, there is still considerable variation for diastase activity itself amongst these lines. The challenge, therefore, is to identify the genes controlling this variation and devise a selection strategy that can be applied by breeders to identify lines with high diastase, and used by official testing authorities and maltsters to promote and stream these varieties for the benefit of the grain distilling market. This will also benefit those spring barley growers that find it challenging to meet the malting specifications for the distilling and brewing markets, as they can safely apply nitrogen to their crops and realise the extra yield benefit whilst obtaining a malting premium for the high diastase market.

A number of studies have looked at the genetic basis of malting quality traits, including DP as well as α-amylase and β-amylase activity, and identified a number of QTL across the barley genome Although no structural diastase genes were identified, a number of genes with a putative role in general carbohydrate metabolism were associated with variation in DP (Marquez-Cedillo et al., 2000; Panozzo et al., 2007; Islamovic et al., 2014). Recently, high density SNP (single-nucleotide polymorphism) genotyping has allowed the use of genome-wide association scans (GWAS) to survey genetic variation influencing malting quality characteristics in diverse collections (Castro et al., 2013). We hypothesise that this type of genome-wide genetic analysis, combined with high resolution sequencing, will identify the SNPs and markers that are contributing to the character. Our approach used pre-existing data and resources gathered under projects, such as Association Genetics of UK Elite Barley (AGOUEB; HGCA Project Report 528) and the BBSRC Crop

Improvement Club Barley Malt Processability project (BB/J019593/1), to identify subsets of germplasm that contrast for diastase activity. These pools were then sequenced using a genome complexity reduction approach, known as exome capture (EC) sequencing that allows all of the known gene space in a barley line to be sequenced easily and cost effectively. Sequencing pooled DNAs, rather than individuals, provides a rapid and highly cost-efficient profile of the quantitative differences in allele frequencies between the two pools. The differential SNPs identified are then used to generate allele-specific KASP markers, which are currently the breeders' marker system of choice, for high-throughput selection. To test whether these markers are good predictors of DP, we validated them by using independent germplasm selected from amongst the remainder of the AGOUEB and IMPROMALT (BB/K008188/1; RD-2012-3776) sets that had already been assessed for DP. In addition, we used the markers to predict the DP of an additional, unknown set of germplasm, supplied by UK and European breeders, which were sown in field trials, and samples from harvested plots were micro-malted and scored for DP in collaboration with the Scotch Whisky Research Institute (SWRI) to test our marker predictions. Finally, we recognised the need to better understand the interaction of DP with different nitrogen fertiliser management. Therefore, using a subset of spring and winter lines, a more detailed phenotyping of DP and its component characters under a range of conditions were undertaken at different field sites. The aim being to differentiate between high and low DP lines so that we can develop and release an improved assessment protocol to complement the markers in an integrated screening package.

The key objectives were:

1. Identify subsets of barley accessions that contrast for diastase activity using pre-existing data and resources, and sequence pools of each subset to identify SNPs that correlate with the contrasts in diastatic power. From these, develop simple easy-to-use markers to differentiate between lines with high and low diastase;
2. Validate these markers as predictors of DP using: a) independent germplasm and; b) a range of germplasm from UK and European breeders which have little or no phenotypic data. The latter set was phenotyped by the Scotch Whisky Research Institute (SWRI) to provide measures of DP that we can use to test our marker predictions;
3. Explore a range of phenotypic screening protocols to develop a robust method that will markedly improve estimation of the genetic potential of an individual line's diastatic power and the enzymes and inhibitors involved, which can then be combined with the marker screening to identify the best of relatively few pre-selected high DP lines.

3. Materials and methods

3.1. Historical data and selection of contrasting subsets

DP data for a collection of 573 UK barley varieties (358 spring and 215 winter) was collected, taken from historical Recommended and National list trial results over the period 1994-2012, as part of the IMPROMALT project (BBSRC: BB/K008188/1). On average, there were 25 sites across both crop types with DP data for each year. Additional DP data was also sourced from field trials of 100 spring and 100 winter barley lines grown under high and low N management regimes for harvest years 2013 and 2104, as part of BBSRC Crop Improvement Research Club funded project: 'Improving the processability of malting barley' (BBSRC: BB/J019593/1). These were micro-malted and the malt analysed by member companies from the Maltsters Association of Great Britain (MAGB). There was considerable overlap between the lines in each project and the total number of lines for which data was available was 602 (371 spring and 231 winter barley lines). Best Linear Unbiased Predictions (BLUPs) of DP for each genotype were calculated, and contrasting sets of high and low DP lines (12 from each) were selected from each of the winter and spring germplasm.

3.2. Exome capture, sequencing and SNP identification

Seedlings from each of the lines selected for the contrasting sets were grown to the three-leaf stage, and 200 mg of leaf material was removed and flash frozen for DNA extraction. Genomic DNA extractions were made using a DNeasy Plant Mini kit (Qiagen) according to the manufacturer's instructions. Extracted DNA was quantified using a Qubit fluorometer (Thermo Fisher) and an equal quantity of genomic DNA was combined for each line in each set (12 lines x 4 sets). Combined sample pools (100 ng each in total) were used for individual exome capture and sequencing. A custom barley exome capture SeqCap EZ library was used throughout, representing approximately 62 Mbp of the barley reference exome (Mascher et al., 2013). DNA library preparation and exome capture was performed using recommended methods in the SeqCap EZ Library SR User's Guide (Nimblegen Roche). The 4 pooled and captured DNA libraries were then sequenced on an Illumina HiSeq 2500 sequencer to generate 100 bp paired end reads in Rapid-Run mode.

The reads were mapped to the barley pseudomolecules genome reference sequence (Martin Mascher et al., 2017; Sebastian Beier et al., 2017) and SNPs identified, as described in detail in the accompanying paper (Looseley et al., 2017).

Allele frequency estimates for each marker within each set were calculated directly from read numbers of each allele. Allele frequency differences (AFD) for each marker were calculated as (the reference allele frequency of the high DP set) minus (the reference allele frequency of the low DP set), giving a possible range between -1 and 1. Genomic locations with an absolute estimated AFD greater than 0.75 were chosen as putatively associated loci. Associations were identified independently in both winter and spring sets. QTL were considered independent if associated

markers were separated by at least 400 Mbp in centromeric regions, or 10 Mbp in non-centromeric regions (corresponding, very approximately, to 10 cM genetic map distance in each case) (Mascher et al., 2017) in which there was no other marker with an AFD > 0.75.

3.3. Marker design and KASP assays

Flanking sequences from the associated markers identified from the exome capture analysis were extracted from the Morex reference genome and used to design KASP assays. These included at least one marker from each QTL location (the most highly differentiated), as well as differentiated SNPs from putative diastase genes collocated with identified QTL. If no SNP with an allele frequency difference greater than 0.75 was present within the putative diastase gene, the criterion was relaxed to an AFD > 0.60. KASP assays were designed from sequence flanking the SNP (extracted from the Morex pseudomolecule sequence) and supplied to LGC Genomics.

3.4. Validation

In order to verify allele frequency estimates derived from the exome capture sequencing analysis, each of the varieties selected for inclusion in the contrasting sets was individually genotyped for each of the KASP assays using an Applied Biosystems Step One Plus Real-Time PCR system according to the manufacturer's instructions.

The robustness of the marker associations identified from the exome capture sequencing was first tested by using each to classify an independent set of 85 spring and 78 winter varieties from the original cultivar set of 602, and test the significance of the BLUPs for their DP. The markers were also used to predict the DP of 61 winter and 89 spring breeders' selections, that were sown in field trials at JHI in autumn 2015 and spring 2016, respectively with a prophylactic fungicide regime and a higher nitrogen fertiliser application (200 and 160 kg/ha total available N for winter and spring barley respectively) than locally appropriate for malting barley. Leaf material was sampled from the field plots of each line, DNA extracted and genotyped with the KASP markers identified under 3.3.

3.5. DP and other malting trait analysis

The plots were harvested in August 2016, dried to 13-14% moisture content and a 250 g cleaned and graded sub-sample sent to SWRI for micro-malting and DP analysis. Micro-malting was carried out in batches of 16 lines, including two controls (recommended high and low varieties for both springs and winters). The following phenotypes were scored: Nitrogen (% DM); Moisture (%); Extract (%); DP (°Lintner & DM); DU (DU & DM) and Friability (%), according to IBD recommended protocols.

3.6. Developing an improved phenotypic assessment of DP

Subsets of winter and spring lines from the assembled high and low pools were used in the initial screening and protocol development (6 high, 6 low, 6 responsive and 2 controls for both spring and winters). Springs were sown on a high fertility soil site near Lanark and winters on the Bush Estate near Edinburgh, both under standard and high nitrogen fertiliser treatments (95 kg & 145 kg N/ha applied in two splits, one in the seedbed and the other a top-dressing at GS23). Grain nitrogen and yield data were gathered for comparisons.

4. Results

4.1. Selection of contrasting pools

Filtering the accessions from the AGOUEB and CIRC Processability projects on consistency of extreme values of DP allowed us to identify 12 accessions corresponding to each of high and low DP for both winter and spring germplasm (Figure 1; Table 1).

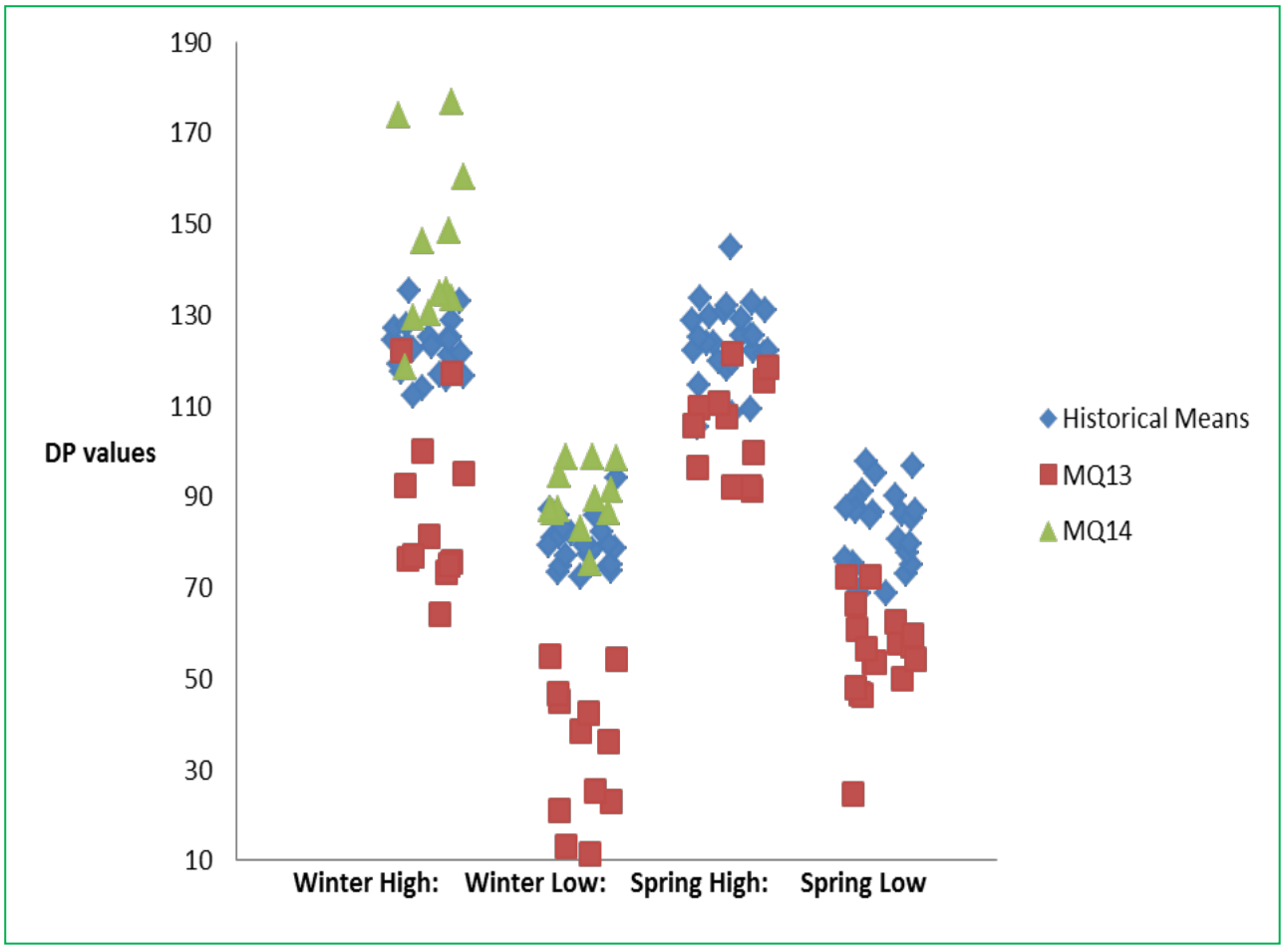

Figure 1. Each coloured shape represents a variety chosen for each of the low and high pools. Blue is the historical mean value over the years it was tested as part of the official trialling process, red and green are its mean DP values from field trials grown for harvest years 2013 and 2014, respectively.

Table 1. Varieties selected for inclusion in each of the contrasting sets.

Spring High DP	Spring Low DP	Winter High DP	Winter Low DP
Belgravia	Alabama	Acute	Cedar
Chime	Amphora	Alpha	Cypress
JB Maltasia	Brazil	Caption	Diadem
Marthe	Cairn	Concept	Diamond
Monika	Calico	Leonie	Fahrenheit
Olympus	Cindy	Melanie	Marinka
Roxana	NSL 95-1257	Milena	Parasol
Sebastian	Otira	Nectaria	Pedigree
Static	Splash	Pearl	Peridot
Tapestry	Spotlight	Silverstone	Portrait
Turnberry	Vivendi	Sunbeam	Prelude
Westminster	Waltz	Torrent	Tallica

4.2. Identification of SNPs between high and low pools for spring and winter lines

Sequencing of contrasting pools for DP generated over 84,000 and 78,000 polymorphic SNPs for spring and winter lines, respectively. From these, we identified 66 markers that distinguished between high and low DP in the winters, corresponding to 6 distinct positions on chromosomes 1H, 2H, 4H and 7H. In the spring germplasm, 32 markers were identified, which located to 3 regions on chromosomes 4H, 5H and 7H (Figure 2). For each region (6 winters and 3 springs), we chose the marker that showed the highest differentiation between pools, several of which were collocated with known diastase genes, or genes which were annotated with a putative diastase function (Figure 2).

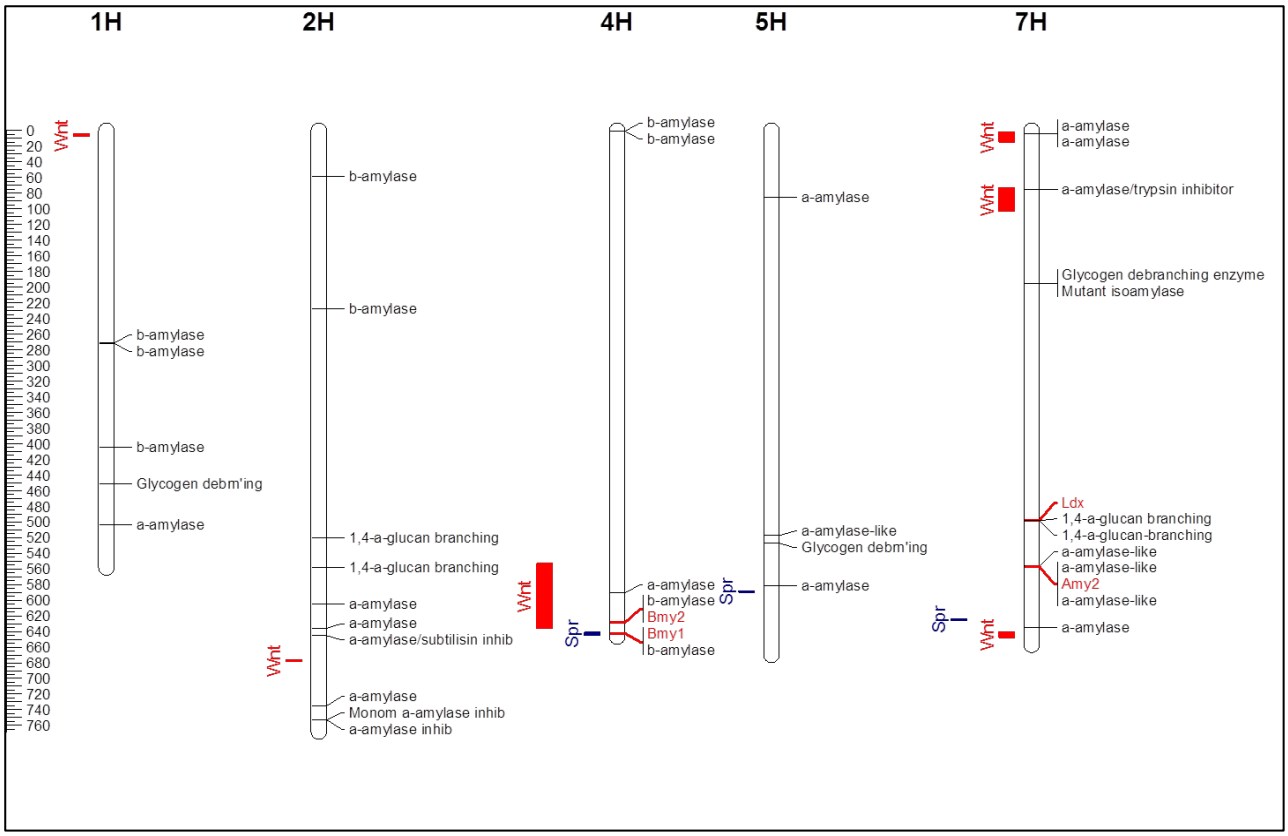

Figure 2. Associations for each contrast are positioned on each of the chromosomes, spring highlighted in blue (Spr) and winters in red (Wnt). Known barley genes associated with diastase activity are also indicated.

4.3. Development of single KASP markers for each of the identified SNPs

Thirteen marker assays were designed from the SNPs identified, 8 for winters (with 5 of these located on chromosome 7H), and 5 for springs (with 2 located on chromosomes 4H and 7H), relevant to annotated diastase genes in these regions (Table 2). We used the KASP assay design which is supplied by LGC genomics and is favoured by UK and European breeders.

Table 2. Details of KASP assays designed from marker associations identified from exome capture sequencing. The population from which the association was identified is indicated along with its chromosome and the name of the gene containing the SNP. Highlighted SNPs were used in the final selection.

Marker	Crop Type	Chromosome	Genes
SNP Assay 1	Winter	1H	12-oxophytodienoate reductase 2
SNP Assay 2	Winter	2H	glutathione peroxidase 6
SNP Assay 3	Winter	4H	Ectonucleoside triphosphate diphosphohydrolase 5
SNP Assay 4	Winter	7H	Glutathione S-transferase family protein
SNP Assay 5	Winter	7H	Acid phosphatase 1
SNP Assay 6	Winter	7H	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein.
SNP Assay 7	Winter	7H	Alpha-amylase/trypsin inhibitor
SNP Assay 8	Winter	7H	unknown
SNP Assay 9	Spring	4H	beta-amylase 5
SNP Assay 10	Spring	4H	Transcriptional coactivator/pterin dehydratase
SNP Assay 11	Spring	5H	unknown function
SNP Assay 12	Spring	7H	unknown
SNP Assay 13	Spring	7H	Alpha-amylase

4.4. Markers validated in extended germplasm with known DP

The selected markers were first validated on individual DNAs from the pooled sampling, 24 winters with 8 markers and 24 springs with 5 markers (highlighted in Table 2). A strong correlation (r = 0.92) between estimated (pooled) and true (individual) was observed for 9 of the 13 SNP assays. In order to test the robustness of the marker associations identified, the independent set of 85 spring and 78 winter varieties was genotyped with the 9 strongly correlated SNP markers. Significant differences between lines carrying high and low marker alleles were observed as shown in Figure 3.

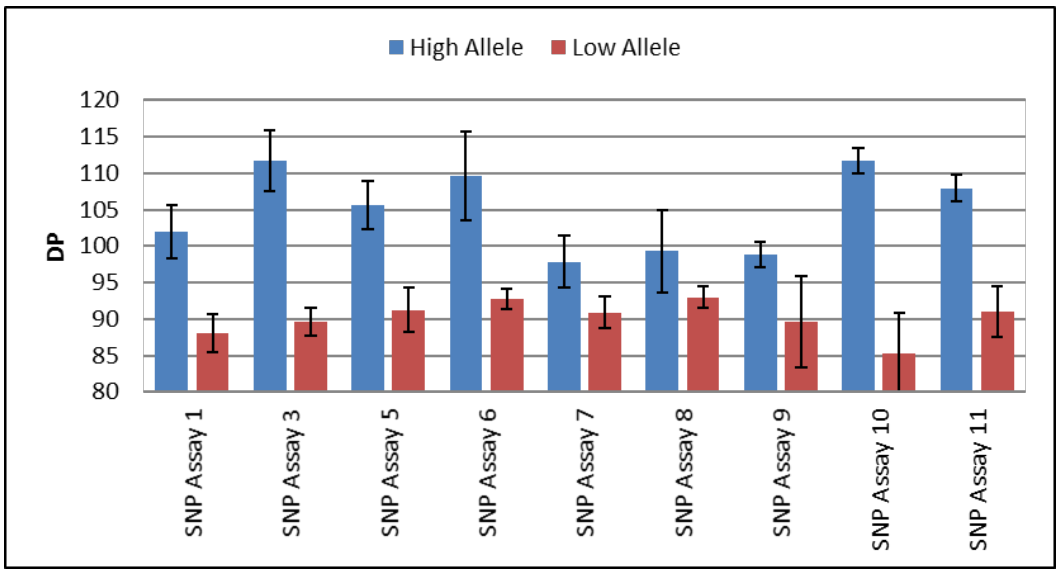

Figure 3. Comparison between DP values of 78 winter and 85 springs with markers scores for each of the 9 KASP markers (SNP Assays 1, 3, 5, 6, 7, 8 winters and 9, 10, 11 springs).

4.5. Markers validated in independent germplasm

In order to further validate the marker, we used the marker profiles to predict whether the lines supplied by the breeders were high or low DP types. Trials were harvested in August 2016 and a subset of 70 lines (28 springs and 42 winters) was chosen on the basis of their predicted DP values. These were assembled in 6 batches of 16, including two spring controls (Concerto and Belgravia) and two winter controls (Flagon and Winsome), with one winter batch representing a full replicate of another that was chosen to maximise the differences in predicted DP. The other two winter batches represented lines predicted to have intermediate DP values. We found that SNP assay 9 was fixed amongst the spring breeders lines, so the two batches were selected on the basis of different combinations of alleles at the two remaining SNP loci. All 96 malting samples were sent to SWRI for micro-malting and subsequent DP analysis. The DP values ranged considerably across batches, from as low as 50 to 197 (°L), with controls also varying from batch to batch. A good indication of reliability is the measure of friability, which in a well-modified malt should be around 85%. However, many of the lines were below this value and ranged from 0.8% to 96.8%, which was indicative of under-modification, which was confirmed by low Soluble Nitrogen Ratios for these malts. Furthermore, malt moisture was also high (> 5%) and this could affect friability scores. Malts with a heavy husk fraction, particularly winter barleys, or malts which have absorbed humidity in the husk layer during storage, may give deviating results in friability measurements. Most samples in the 4 winter batches had extremely low friability, indicating they were under-modified, which explains the very low DU values (α-amylase becomes active during germination; Farzaneh et al., 2017).

Using the lines with high friability, we were able to determine whether using the SNP markers identified high and low DP lines (Table 3; Figure 4). Overall, there was a much better and significant relationship between predicted and actual values for the winter lines (R^2= 0.48) than the springs, which showed little or no correlation. The most likely explanation is that there is now very little variation genetically between the springs, as witnessed by the fixation of the high allele on one of the three predictive SNPs amongst the breeders selections. There was also a highly significant excess of the high allele for both the other SNPs. There was little evidence of fixation amongst the winters, with only 1 of the 5 SNPs showing a significant excess of high alleles. In fact, one of the SNPs (Assay 8) had a highly significant excess of low alleles, with the other three being in equilibrium. This results in a wider range of potential variation amongst the winters, which is borne out by a comparison of the observed phenotypes for the better modified malts, with DP ranging from 103 to 150 for the springs and from 129 to 196 for the winters.

Table 3. Predicted DP, actual DP and other malting traits values.

	DP_dm	Predicted DP	DU_dm	Extract	Friability_%	Moisture
Belgravia	138		73	83.1	89.7	7.7
Concerto	125		69	82.7	95.3	7.6
Flagon	147		33	82.1	57.7	7.5
sdp17	126	90	61	82.8	94.5	7.8
sdp18	150	102	58	82.7	93.7	7.8
sdp19	137	102	42	83.3	93.8	7.6
sdp3	144	84	76	83.5	95.7	8.1
sdp37	134	102	50	83.6	91.1	7.8
sdp39	151	84	47	83.1	94.9	7.8
sdp4	127	102	54	83.8	92.9	7.8
sdp43	104	90	44	82.7	92.9	8.0
sdp53	110	102	53	83.3	89.1	7.6
sdp68	129	90	52	83.4	90.8	7.8
sdp71	151	90	66	82.9	95.2	7.8
sdp72	144	102	60	82.9	94.3	7.7
sdp79	141	84	51	83.6	91.1	7.8
sdp83	107	102	66	82.1	94.5	7.4
wdp1	158	93	42	80.9	49.1	5.9
wdp10	184	118	36	82.8	55.0	7.3
wdp19	141	93	26	81.5	37.9	8.1
wdp21	167	127	34	81.6	60.9	7.4
wdp22	130	93	36	81.8	59.6	7.4
wdp24	132	116	22	83.3	54.5	7.3
wdp25	196	127	29	81.1	59.6	7.2
wdp26	130	93	25	80.3	59.0	8.3
wdp29	154	98	21	81.6	51.9	7.8
wdp30	144	101	21	81.7	44.7	7.5
wdp31	130	110	23	82.6	38.9	8.2
wdp32	180	123	30	82.3	53.1	7.6
wdp33	133	81	23	81.7	36.0	7.8
wdp35	161		29	82.2	54.6	7.3
wdp36	122		26	81.4	37.9	7.5
wdp39	211		39	82.2	70.6	8.2

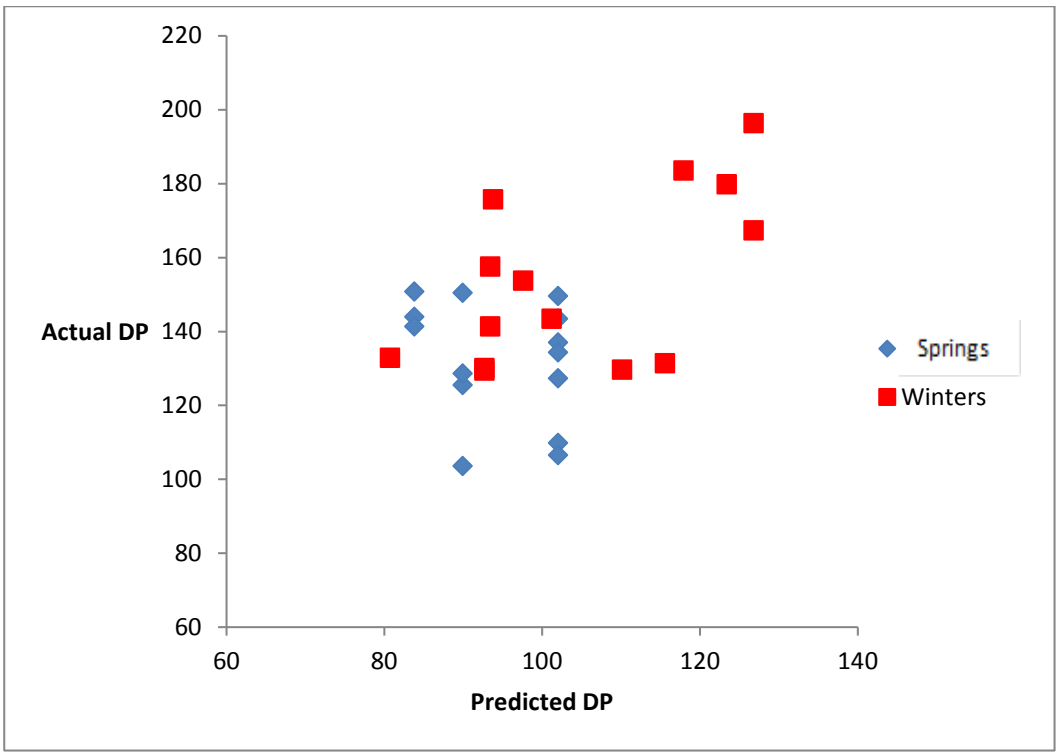

Figure 4. Relationship between predicted and actual DP values, for spring (blue diamonds) and winters (red squares).

4.6. Development of procedures for phenotype screens

Grain nitrogen and yield data were gathered from the Lanark and Bush Estate trials and, in general, there was no yield increase due to elevated nitrogen, and very little difference in yield based on DP designation (high, low or responsive). In contrast, grain nitrogen differed between subsets; at the higher nitrogen rate from 2.02 to 1.84%, and the lower rate from 1.60 to 1.47%, for the high DP lines and low DP lines, respectively. Fig. 5 shows the relationship between grain N% and yield for the different high, low and responsive lines.

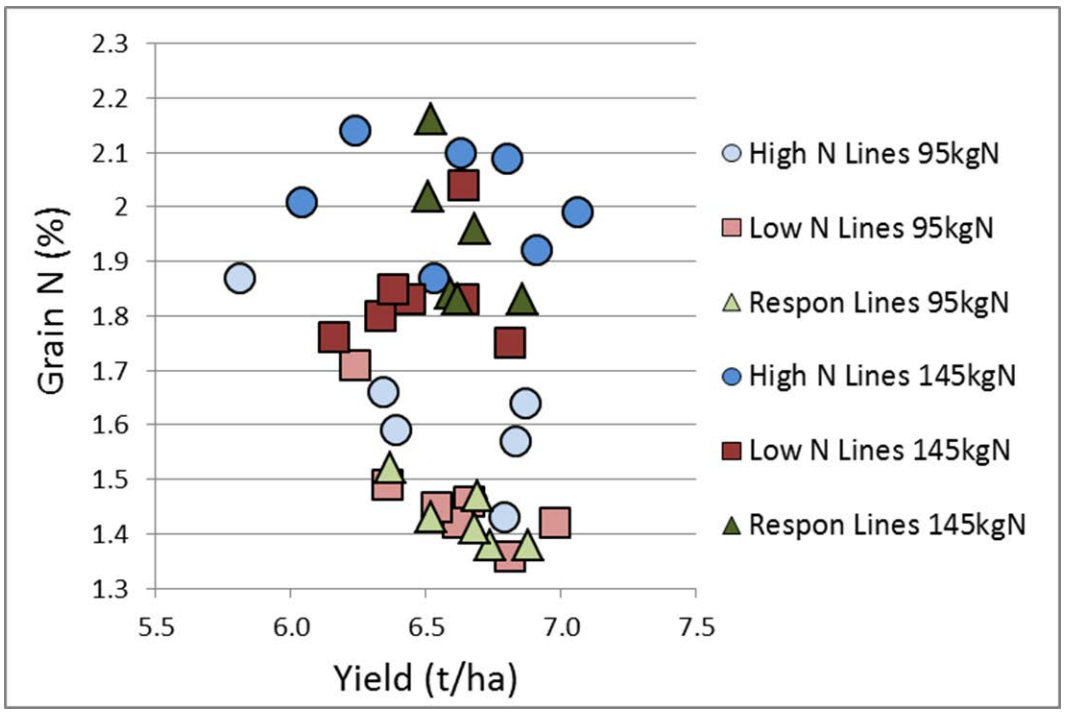

Fig. 5. Relationship between grain N% and yield for high, low and responsive barley lines grown at total N fertiliser applied of 95 or 145 kg N ha^{-1}.

At the lower N supply of 95 kg N ha^{-1}, the high N lines tended to have an increased grain N concentration compared to the low and responsive N lines, whilst at the high N supply of 145 kg N ha^{-1}, both high N and responsive N lines had relatively high grain N concentration. To investigate further, a more detailed field trial was sown in 2016 at 3 sites, under 7 nitrogen regimes and with 4 spring barley varieties. The results are highlighted in Table 4, indicating that grain yield did not change significantly with delayed N fertiliser application, but grain N concentration increased when the split in N fertiliser was applied at later early growth stages, e.g. stem extension (GS31). Varieties responded differently to the N fertiliser splits. With the exception of the three-way N split, the high yielding and low N variety Laureate had grain N concentration below that of varieties Belgravia, Olympus and Westminster.

Table 4. Yield *(t/Ha)* and grain nitrogen (%) data for Bush estate, under 7 nitrogen regimes and with 4 varieties. Numbers in brackets in the fertiliser column indicate the growth stage of the top dressing following an application of 60 kg N/ha in the seedbed.

Fertiliser	Belgravia	Laureate	Olympus	Westminster
0	4.95	5.28	5.56	4.61
60 + 60 (13)	6.7	7.49	7.16	6.82
60 (seed bed only)	6.30	7.07	7.54	6.57
60 + 60 (21)	6.81	8.04	7.20	6.59
60 + 60 (31)	6.78	7.89	7.16	6.76
60 + 60 (37)	6.88	7.58	7.11	6.74
60 + 30 (21) + 30 (37)	6.79	6.60	7.28	6.43

Fertiliser	Belgravia	Laureate	Olympus	Westminster
0	1.23	1.16	1.21	1.33
60 + 60 (13)	1.87	1.75	1.78	1.72
60 (seed bed only)	1.73	1.50	1.68	1.70
60 + 60 (21)	1.78	1.76	1.68	1.81
60 + 60 (31)	1.83	1.65	1.83	2.00
60 + 60 (37)	1.83	1.63	1.73	1.84
60 + 30 (21) + 30 (37)	1.72	1.91	1.77	1.93

Discussion

The results presented here demonstrate the effectiveness of using allele frequency estimates from pooled high density sequencing data of phenotypically contrasting barley sets to map quantitative traits. This approach was used to identify genetic markers associated with variation in a complex malting quality trait, which were subsequently confirmed by genotyping using a separate set of lines.

4.7. Diastase genes & associations

The majority of the associated loci identified were from winter barley, suggesting that this germplasm set is more diverse for genes influencing DP. This is consistent with a higher genetic diversity in winter germplasm in general (Thomas et al., 2014), but may also reflect historical selection for malting quality traits in spring barley, effectively resulting in the near fixation at malting quality loci. This hypothesis is supported by the observation of higher variability in DP estimates seen in winter varieties, compared to springs.

The availability of genome pseudomolecule sequences for each of the barley chromosomes now allows QTL regions to easily be placed in the context of the underlying gene content. Homology based searching of the reference genome identified a number of genes that may be diastase related (see Figure 2). These included several α- and β-amylases located in genomic regions not previously reported to be associated with diastase activity. The majority of QTL identified aligned with genes annotated as diastase-related. Whilst this study does not present unequivocal evidence to suggest that these genes are causal to the identified QTL, the correspondence between these and the positions of the QTL provides a set of potential gene candidates, as well as offering further support to the identification of these regions as influencing QTL affecting DP. The generation of sequence data from high and low DP varieties across these loci will help to address whether variation at these genes is responsible for differences in DP seen within the cultivar collection.

In both winter and spring contrasts, associations were detected in the telomeric region of chromosome 4HL. This region has previously been associated with QTL for DP and β–amylase activity, and contains two known β–amylase genes (*Bmy1* & *Bmy3*). Previous studies have linked variation at the *Bmy1* locus with variation in DP (Hayes et al., 1997; Coventry et al., 2003). Results presented in this study suggest that an additional α-amylase gene (HORVU4Hr1G073630) may also be present in this region, although no differentiated SNPs were identified within this gene from the sequence data generated here. Variation at these genes may be associated with differences in DP in both spring and winter barley and, as such, they represent good candidates for detailed characterisation. The improved marker density of genome-wide approaches, when combined with exome capture sequencing and the new genome sequence, offer the potential to increase the resolution of mapped QTL. Indeed, the results presented here suggest that variation at the *Bmy1*

locus is unlikely to be responsible for the QTL on chromosome 4H in winter varieties, but is more likely to be responsible for the QTL seen in the spring contrast. Nevertheless, the SNP marker designed in the *Bmy1* gene (SNP assay 9) was not as highly associated with DP as another closely linked marker (SNP Assay 10). This observation could either be interpreted as reflecting a gene (linked to *Bmy1*) influencing the trait, or being caused by genetic diversity within a linkage block containing both markers; causing a lower correlation between the causal variant and *Bmy1* SNP. There are 97 reported high confidence gene models across the QTL interval (Mascher et al., 2017), some of which may also influence carbohydrate metabolism from their annotation. Detailed characterisation of each of the candidate genes in the variety collection may help to address the specific genetic control of DP across this region of chromosome 4H, which would be of considerable use for breeders in both the selection of parental lines for crosses, as well as in marker assisted selection.

The majority of QTL identified collocate with known (or putative) structural diastase genes. One of the candidate QTL (at 75 Mbp on chromosome 7H) is linked to a putative α-amylase inhibitor (HORVU7Hr1G035020), suggesting that genetic variation at endogenous amylase inhibitor loci may also influence malt diastase activity. Indeed, a SNP assay designed in this gene showed a stronger association with DP in the wider variety collection than the peak marker identified from the exome capture data. As such, this represents an extremely strong candidate for follow-up studies. Variation in expression of an α-amylase/subtilisin inhibitor has previously been described as correlating with β-amylase activity (Potokina et al., 2004), but the expressed sequence tag (EST) identified by that study (HY06J10V) maps to the telomeric region of chromosome 2H on the Morex reference assembly, distal to the QTL identified from the winter contrast in this work and is located in gene HORVU2Hr1G090750 (Figure 2).

One of the strongest associations identified in the current study was in the telomeric region of the short arm of chromosome 1H. This is potentially collocated with a QTL reported in a contrast between Australian and Canadian malting barleys (Zhou et al., 2016). The 3.8 Mbp interval identified here contains 121 high confidence genes (Mascher et al., 2017). Whilst none of these are annotated as having a putative diastase function, a large number are annotated as having a role in protein or carbohydrate metabolism. As such, detailed characterisation of this region will be required to identify candidate genes underlying this QTL effect.

The putative QTL effect on chromosome 2H identified in the winter contrast was only supported by one differentiated marker, and individual genotyping of the selected lines showed that this was likely to be an overestimation and thus represents a spurious association. This illustrates the importance of identifying multiple differentiated markers to avoid false associations when using estimates of allele frequencies for association mapping.

4.8. Allele frequency estimates from pooled DNAs

Central to the mapping method employed in this study is the ability to accurately estimate allele frequencies from next generation sequencing of pooled DNA samples. A number of potential sources of error might influence estimates of allele frequencies, including uneven contribution of individuals, preferential capture of alleles, or sampling errors when read coverage is low (Gautier et al., 2013). Whilst experimental sources of error (such as uneven contribution of individual lines to pooled samples) can only be controlled by careful laboratory technique, errors associated with read sampling can be adjusted during analysis. Individual genotyping of lines from each contrasting set confirmed that the majority of the markers, that were identified as being highly differentiated from pooled exome capture reads, were genuinely differentiated between high and low sets. This suggests that, when appropriate filtering of variant calls is conducted, allele frequency estimates from mixed sample NGS data sets offer the ability to accurately identify differences between combined DNA samples. This allows considerable savings in cost, as well as the ability to sample higher numbers of lines than would be possible if individual libraries had to be prepared.

4.9. Validation of associations

A further question addressed by this study was whether differentiation between phenotypically contrasting sets was an effective method for identifying genuine marker trait associations. To address this, associations were tested by genotyping in a wider germplasm collection in order to allow an independent validation of candidate markers. This analysis showed that significant differences (in historical DP estimates) were present between lines carrying alternative alleles at candidate markers. However, when validation was carried out on a set of breeding lines with unknown DP values, the results were not so clear. It is likely that there was some residual dormancy amongst the material, as the samples were sent for malting without adequate allowance for natural dormancy break due to the confines of the project timescale. This, together with the potential for moisture accumulation in the husk during storage, will result in under-modified malt, as indicated by some low SNR and friability values, and hence under-expression of DP. From the data obtained by filtering out the worst malts, we did see a highly significant positive relationship between predicted DP and the observed DP for the winter lines. In addition, it was most noticeable that the levels of DP were much higher amongst the winters, indicating that the winter crop would be better suited to the production of high DP barley and that the markers we have identified could successfully be used in a selection programme. By contrast, there appears to be little prospect for much improvement in DP amongst the current elite spring gene pool and further progress would require introgression of high DP from other gene pools, such as the North American 6-row elite material. It remains to be seen whether the markers we identified for the UK elite springs would be valuable in selection in crossing with such material, and further research would be needed to

devise a suitable strategy to introgress DP from another gene pool. Ideally, further research should be carried out to test the value of the markers in prediction more extensively and also across introgression programmes.

In conclusion, this study demonstrates that the use of new high resolution sequencing of pooled DNA from phenotypically contrasting sets represents a powerful method for conducting genome-wide association studies to identify a number of novel QTL for the genetic control of DP in UK barley, and provided candidate genes that can be followed up in subsequent studies. We recommend that the markers that we have identified be used in a strategy to develop high DP winter barley for use in the UK.

5. References

Beier, S., Himmelbach, A., Colmsee, C., et al. (2017). Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. Scientific Data 4, 170044. doi: 10.1038/sdata.2017.44

Castro, A., Cammarota, L., Gomez, B., Gutierrez, L., Hayes, P.M., Locatelli, A., et al. (2013). "Genome-Wide Association Mapping of Malting Quality Traits in Relevant Barley Germplasm in Uruguay," in Advance in Barley Sciences: Proceedings of 11th International Barley Genetics Symposium, eds. G. Zhang, C. Li & X. Liu. (Dordrecht: Springer Netherlands), 37-46.

Coventry, S.J., Collins, H.M., Barr, A.R., Jefferies, S.P., Chalmers, K.J., Logue, S.J., et al. (2003). Use of putative QTLs and structural genes in marker assisted selection for diastatic power in malting barley (*Hordeum vulgare* L.). Australian Journal of Agricultural Research 54(11-12), 1241-1250. doi: 10.1071/Ar02193.

Farzaneh, V., Ghodsvali, A., Bakhshabadi, H., Zare, Z., Carvalho, IS (2017). The impact of germination time on some selected parameters through malting process. International Journal of Biological Macromolecules 94, 663-668.

Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., et al. (2013). Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. Molecular Ecology 22(14), 3766-3779. doi: 10.1111/mec.12360.

Hayes, P.M., Cerono, J., Witsenhoer, H., Kuiper, M., Zabeau, M., Sato, K., et al. (1997). Characterizing and exploiting genetic diversity and quantitative traits in barley (*Hordeum vulgare*) using AFLP markers. Journal of Agricultural Genomics 3, 1-18.

Islamovic, E., Obert, D.E., Budde, A.D., Schmitt, M., Brunick, R., Kilian, A., et al. (2014). Quantitative trait loci of barley malting quality trait components in the Stellar/01Ab8219 mapping population. Molecular Breeding 34(1), 59-73. doi: 10.1007/s11032-014-0017-3.

Looseley, M., Bayer, M., Bull, H., Ramsay, L., Thomas, W., Booth, A., De La Fuente Canto, C., Morris, J., Hedley, P. and Russell, J. (2017). Association mapping of diastatic power in UK winter and spring barley cultivar collections by exome capture sequencing of contrasting variety sets. Frontiers in Plant Science (submitted).

Marquez-Cedillo, L.A., Hayes, P.M., Jones, B.L., Kleinhofs, A., Legge, W.G., Rossnagel, B.G., et al. (2000). QTL analysis of malting quality in barley based on the doubled-haploid progeny of two

elite North American varieties representing different germplasm groups. Theoretical and Applied Genetics 101(1), 173-184. doi: 10.1007/s001220051466.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. Nature 544(7651), 427-433

Mascher, M., Richmond, T.A., Gerhardt, D.J., Himmelbach, A., Clissold, L., Sampath, D., et al. (2013). Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. The Plant Journal 76(3), 494-505. doi: 10.1111/tpj.12294.

Panozzo, J.F., Eckermann, P.J., Mather, D.E., Moody, D.B., Black, C.K., Collins, H.M., et al. (2007). QTL analysis of malting quality traits in two barley populations. Australian Journal of Agricultural Research 58(9), 858-866. doi: https://doi.org/10.1071/AR06203.

Potokina, E., Caspers, M., Prasad, M., Kota, R., Zhang, H., Sreenivasulu, N., et al. (2004). Functional association between malting quality trait components and cDNA array based expression patterns in barley (*Hordeum vulgare L*.). Molecular Breeding 14(2), 153-170. doi: 10.1023/B:MOLB.0000038004.21342.3c.

Thomas, B., Comadran, J., Ramsay, L., Shaw, P., Marshall, D., Newton, A., et al. (2014). "HGCA Project Report 528: Association genetics of UK elite barley (AGOUEB)".).

Zhou, G.F., Panozzo, J., Zhang, X.Q., Cakir, M., Harasymow, S., and Li, C.D. (2016). QTL mapping reveals genetic architectures of malting quality between Australian and Canadian malting barley (*Hordeum vulgare* L.). Molecular breeding v. 36(no. 6), pp. 70-70-2016 v.2036 no.2016. doi: 10.1007/s11032-016-0492-9.